



10/506504

Rec'd PCT/PTO 03 SEP 2004



INVESTOR IN PEOPLE

The Patent Office  
 Concept House  
 Cardiff Road  
 Newport  
 South Wales  
 NP10

REQD 23 APR 2003

WIPO

PCT

**PRIORITY DOCUMENT**  
 SUBMITTED OR TRANSMITTED IN  
 COMPLIANCE WITH  
 RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

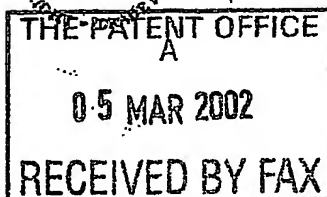
Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 7 April 2003

**BEST AVAILABLE COPY**

## Patents Form 1/77

Patents Act 1977  
(Rule 16)05HAR02 E700861-1 002987  
P01/7700 0-00-0205097.9

## Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form.)

The Patent Office

Cardiff Road  
Newport  
South Wales  
NP10 8QQ

1. Your reference

2522

2. Patent application number

(The Patent Office will fill in this part)

0205097.9

3. Full name, address and postcode of the or of each applicant (underline all surnames)

UNIVERSITY OF SOUTHAMPTON  
HIGHFIELD  
SOUTHAMPTON  
SO17 1BJ

Patents ADP number (if you know it)

If the applicant is a corporate body, give the country/state of its incorporation

UNITED KINGDOM 798470001

4. Title of the invention

NATURAL LANGUAGE PROCESSING FOR EXPERTISE  
MODELLING IN E-MAIL COMMUNICATION.

5. Name of your agent (if you have one)

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

CENTRE FOR ENTERPRISE AND INNOVATION  
UNIVERSITY OF SOUTHAMPTON  
HIGHFIELD  
SOUTHAMPTON  
SO17 1BJ.

Patents ADP number (if you know it)

8019325001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number  
(if you know it)Date of filing  
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing  
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if

a) any applicant named in part 3 is not an inventor, or  
b) there is an inventor who is not named as an applicant, or

c) any named applicant is a corporate body.

See note (d)

yes

Patents Form 1/77

0030478 05-Mar-02 12:10

## Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 6 ✓ *Jim*Claim(s) —Abstract —Drawing(s) —

10. If you are also filing any of the following, state how many against each item.

Priority documents —Translations of priority documents —Statement of inventorship and right to grant of a patent (Patents Form 7/77) —Request for preliminary examination and search (Patents Form 9/77) —Request for substantive examination (Patents Form 10/77) —Any other documents (please specify) —

11.

I/We request the grant of a patent on the basis of this application.

Signature *Rayner*Date 5/3/2002

12. Name and daytime telephone number of person to contact in the United Kingdom

MYLENE ROYAERT 023 80594680**Warning**

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

**Notes**

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- e) Once you have filled in the form you must remember to sign and date it.
- f) For details of the fee and ways to pay please contact the Patent Office.

Patents Form 1/77

D030478 05-Mar-02 12:10

## Natural Language Processing for Expertise Modelling in E-mail Communication\*

**Abstract.** One way to find the information that may be required, is to approach a person who is believed to possess it or to identify a person who knows where to look for it. Technical support, which automatically compiles individual expertise and makes this accessible, may be centred on an expert finder system. A central component of such a system is a user profile, which describes user expertise level in discussed subjects. Previous works have made attempts to weight user expertise by using content-based methods, which associate the expertise level with the analysis of keyword usage, irrespective of any semantic meanings conveyed. This paper explores the idea of using a natural language processing technique to understand given information from both a structural and semantic perspective in building user profiles. With its improved interpretation capability compared to prior works, it aims to enhance the performance accuracy in ranking the order of names of experts, returned by a system against a help-seeking query. To demonstrate its efficiency, e-mail communication is chosen as an application domain, since its closeness to a spoken dialog, makes it possible to focus on the linguistic attributes of user information in the process of expertise modelling. Experimental results from a case study show a 23% higher performance on average over 77% of the queries tested with the approach presented here.

### 1. Introduction

A crucial task in the distributed environments that most organizations operate is to effectively manage the useful knowledge held by individuals. Not only does this supplement additional resources, but it also contributes timely and up-to-date procedural and factual knowledge to enterprises. In order to fully maximize individually held resources, it is necessary to encourage people to share such valuable data. As their expertise is accumulated through task achievement, it is also important to exploit it as it is created. Such an approach allows individuals to work as normal without demanding changes in working environments [7].

An expert finder is a system designed to locate people who have 'bought-after knowledge' to solve a specific problem [4]. It answers with the names of potential helpers against knowledge seeking queries, in order to establish personal contacts which link novices to experts. The ultimate goal of such a system is to create

---

\* This work was funded by the University Technology Partnership (UTP) for Design, which is a collaboration between Rolls-Royce, BAE Systems and the Universities of Cambridge, Sheffield and Southampton.

environments where users are aware of each other, maximizing their current resources and actively exchanging up-to-date information. Although the expert finder systems cannot always generate correct answers, bringing the relevant people together provides opportunities for them to become aware of each other, and to have further discussions, which may uncover hidden expertise.

In designing technical support to maximize the use of such personal expertise, two issues have to be addressed; 1) how to simulate the personal contacts found in real environments, and 2) how to capture personal expertise while allowing users to work as they normally do without demanding changes in working environments. The exploitation of e-mail communication, which can be enhanced as a communication-based learning tool, where individual experiences are shared among communicators, is proposed as an enabling technology. E-mail communication has become a major means of exchanging information and acquiring social or organisational relationships, implying that it would be a good source of information about recent and useful cooperative activities among users. It is hypothesized that because of its popularity, information mined from e-mail communication can be considered as information from expertise discovery sources [3; 7]. In addition, as it represents an every day activity, it requires no major changes to working environments, which makes it suitable as a test environment.

A decision about whether an individual is an expert for a given problem may be made by consulting user profiles. Drawn from information retrieval studies, the frequencies of keywords have been extensively used for extracting user information from exchanged e-mail messages. However, there are at least three reasons why such an approach is inadequate when applied to expertise modelling. First, counting keywords is not adequate for determining whether a given document is factual information or contains some level of author expertise. Secondly, without understanding the semantic meanings of keywords, it is possible to assume that different words represent the same concept and vice versa, which triggers the retrieval of non-relevant information. Finally, it is not easy to distinguish question-type texts from potential answer documents, which support retrieval of the relevant documents for the given query. In addition, the argument that user expertise is action-centred and is often distributed in the individual's action-experiences, is the motivation behind this work that relies on linguistic-oriented user modelling [2]. With this approach, when we regard the given messages as the realization of involved knowledge, user expertise can be verbalized as a direct indication of user views on discussed subjects, and the levels of expertise are distinguished by taking into account the degree of significance of the words employed in the messages.

In this paper, a new expertise model, EMNLP (Expertise Modelling using Natural Language Processing) that captures the different levels of expertise reflected in exchanged e-mail messages, and makes use of such expertise in facilitating a correct ranking of experts, is presented. It examines the application of NLP (Natural Language Processing) techniques and user modelling to the development of an expert finder system based on e-mail communication. The creation of an expert finder system that promptly accommodates new information is one of the two main themes of this paper, and while improving its competency values by using NLP for profiling of users is the second.

## 2 Related Work

KnowledgeMail from Tacit Corp is the system most related to EMNLP, in that it adds an automatic profiling ability to some of existing commercial e-mail systems, to support information sharing through executing queries about the profiles constructed [7]. User profiles are formulated as a list of weight-valued terms by using one statistical method. A survey focusing on the system's performance reveals that users tend to spend extra time cleaning up their profiles in order to reduce false hits, which erroneously recommend them as experts due to unresolved ambiguous terms [3]. In an effort to reduce such problems, the application of NLP to profiling users is suggested. As a consequence, EMNLP is expected to generate more meaningful terms in user profiles.

Maybury et al. [4] developed an expert finder system that exploits the intellectual products created within an organisation to support automated expertise identification. The system considered a user as an expert if he/she was linked to a wide range of documents and/or a large number of documents about that topic. It combines multiple evidence demonstrating associations with the user in determining the level of expertise of the user. This is comparable to EMNLP in that it qualifies experts by requiring detailed evidence, however, it differs in that such evidence is collected from the measurement of information usage patterns, rather than from the analysis of the meanings and functional roles of such information.

Based on the Java Programming domain, the system described by Vivacqua et al. [9] model a user's programming skill by reading source code files, and analysing what classes, libraries or methods are used and how often. This result is then compared to the overall usage for the remaining users, to determine the levels of expertise for specific topics (e.g., methods). Its automatic profiling and mapping of five levels of expertise (i.e., expert-advanced-intermediate-beginner-novice) are similar to those of EMNLP. However, the expertise assignment function is rather too simplified in so far as it disregards various coding patterns that might reveal the different skills of experts and beginners.

## 3 Descriptions of EMNLP

A design objective of EMNLP is to improve the efficiency of the task search, which ranks peoples' names in decreasing order of expertise against a help-seeking query. Its contribution is to turn once simply archived e-mail messages into knowledge repositories by approaching them from a linguistic perspective, which regards the exchanged messages as the realization of verbal communication among users. Its supporting assumption is that user expertise is best extracted by focusing on the sentence where users' viewpoints are explicitly expressed. NLP is identified as an enabling technology that analyzes e-mail messages with two aims: 1) to classify sentences into syntactical structures (syntactic analysis), and 2) to extract users' expertise levels using the functional roles of given sentences (semantic interpretation). Figure 1 shows the procedure for using EMNLP, i.e. how to create user profiles from the collected messages. Further details of the NLP components are explained with the dotted line. Contents are decomposed into a set of paragraphs and heuristics (e.g., locating a full stop) are applied in order to break down each paragraph into sentences.

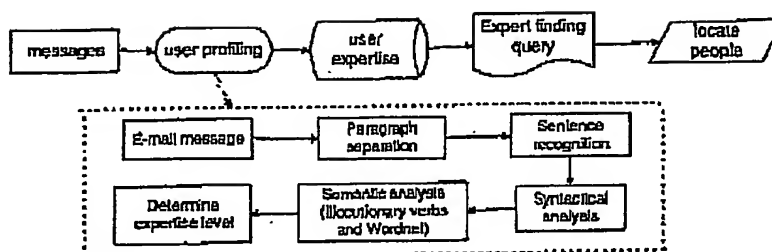


Fig.1: The procedure for using an NLP-based user profiling

Syntactical analysis identifies the syntactic roles of words in a sentence by using a corpus annotation [1]. Apple Pie Parser, which is a bottom-up probabilistic chart parser and finds the parse tree with the best score by the best-first search algorithm, is used for this purpose [6]. The syntactical analysis supports the location of a main verb in a sentence, by decomposing the sentence into a group of grammatically related phrases, such as "noun", "adverb", "adjective", "verb", or "preposition".

Given the structural information about each sentence, semantic analysis examines sentences with two criteria: 1) whether the employed verb verbalizes the speaker's attitudes, and 2) whether the sentence has a "first person" (e.g., "I", "In my opinion", or "We") subject. This analysis is based on Speech Act Theory (SAT), which proposes that communication involves the speaker's expression of an attitude (i.e. an illocutionary act) towards the contents of the communication [8]. It suggests that information can be delivered with different communication effects on recipients depending on different speaker's attitudes, which are expressed using an appropriate illocutionary act, which represents a particular function of communication. The performance of the speech act is described by a verb, which posits a core element as the central organizer of the sentence. In addition, the fact that working practices are reflected through task achievement implies that personal expertise can be regarded as action-oriented, emphasizing the important role of a "first person" subject in expertise modelling [2].

EMNLP extracts user expertise from the sentences, which have "first person" subjects, and determines expertise levels based on the identified main verbs. Whereas SAT reasons about how different illocutionary verbs convey the various intentions of speakers, NLP determines the intention by mapping the central verb in the sentence to the pre-defined illocutionary verb. The decision about the level of user expertise is made according to the defined hierarchies of the verbs, initially provided by SAT. SAT provides the categories of illocutionary verbs (i.e. assertive, commissive, directive, declarative, and expressive), each of which contains a set of exemplary verbs. EMNLP further extends the hierarchy in order to increase its coverage for practicability by using the WordNet Database [5]. EMNLP first examines all verbs occurring in the collected messages, and then filters out verbs, which have not been mapped onto the hierarchy. For each verb, it consults the WordNet database in order to assign a value through chaining its synonyms; for example, if the synonym of the given verb is classified into "assertive" value, and then this verb is also assigned into "assertive".

To clarify how two sentences, which may be assumed to contain similar keywords, are mapped onto different profiles, consider two example sentences: 1) "For the 5049 testing, phase analysis on those high frequency results that Rob plotted is needed", and 2) "For the 5049 testing, I know we need phase analysis on those high frequency results that Rob plotted". The main verb values for both sentences (i.e., need and know) are equivalent to "Strong Working Knowledge", which conveys a relatively high knowledge for a speaker. However, the difference is that when compared to the first, the second sentence clearly conveys the speaker's intention as it begins with "I know". As a consequence, it is regarded as demonstrating expertise while the first sentence is not. Information extracted from the first sentence is mapped onto a lower-level expertise.

#### 4 Experimental Results

A case study has been developed to test two hypotheses: namely 1) that EMNLP produces comparable or higher accuracy in differentiating expertise from factual information compared to that of the frequency-based statistical model, 2) that differentiating expertise from factual information supports more effective query processing in locating the right experts. As a baseline, a frequency-based statistical model, which builds user profiles by weighting presented terms without considering their meanings or purposes was used.

A total of 10 users, who work for the same department in a professional engineering design company, participated in the experiment and a period of three-to-four months duration was spent collecting e-mail messages. A total of 18 queries were created for a testing dataset, and a maximum number of 40 names of predicted experts, i.e. 20 names extracted using EMNLP and 20 names from the statistical model, were shown to a user, who was the group leader of the other users. As a manager, the user was able to evaluate the retrieved names according to the five pre-defined expertise levels: "Expert-Level Knowledge", "Strong Working Knowledge", "Working Knowledge", "Strong Working Interests", and "Working Interests".

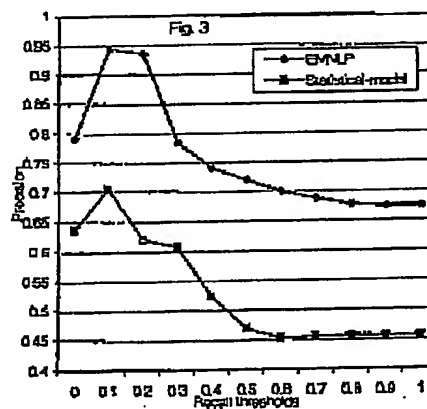
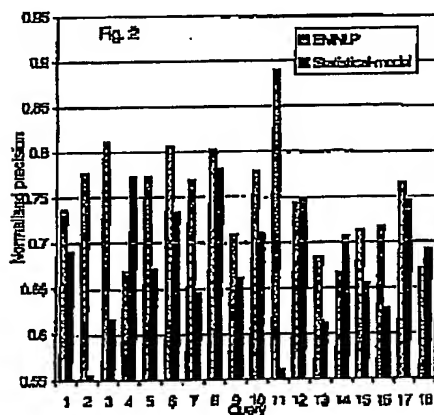




Figure 2 summarizes the results measured by normalised precision. For 4 questions (i.e. 4,12,14,18), EMNLP produced lower performance rates than by using the statistical approach. However, for 14 queries, its ranking results were more accurate, and at the highest point, it outperformed the statistical method with a 33% higher precision value. The precision-recall curve, which demonstrates a 23% higher precision value for EMNLP, is shown in Figure 3. The differences of precision values at different recall thresholds are rather small with EMNLP, implying that its precision values are relatively higher than those of the statistical model.

A close examination of the queries used for testing reveals that the statistical model has a better capability in processing general-type queries that search for non-specific factual information, since 1) as we regard user expertise as action-oriented, knowledge is distinguished from such factual information, implying that it is difficult to value factual information as knowledge with EMNLP, and 2) EMNLP is limited to exploring various ways of determining the level of expertise in that it constrains user expertise to be expressed through the first person in a sentence.

## 5. Future Work

EMNLP was developed to improve the accuracy of ranking the order of expert names by use of the NLP technique to capture explicitly stated user expertise, which otherwise may be ignored. Its improved ranking order, compared to that of a statistical method, was mainly due to the use of an enriched expertise acquisition technique, which successfully distinguished experienced users from novices. We presume that EMNLP would be particularly useful when applied to large organizations where it is vital to improve retrieval performance since typical queries may be answered with a list of a few hundred potential expert names.

Special attention is given to gathering domain specific terminologies possibly collected from technical documents such as task manuals or memos. This is particularly useful for the semantic analysis, which identifies concepts and relationships within the NLP framework, since these terminologies are not retrievable from general-purpose dictionaries (e.g. the WordNet database).

## References

1. Allen, J. (1987) *Natural Language Understanding*, Benjamin/Cummings Publishing
2. Choo, C. W., Detlor, B., Turnbull, D. (2000) *WEB WORK Information Seeking and Knowledge Work on the World Wide Web*, Kluwer Academic Publishers
3. Forbes (2001) Forbes, You've got expertise, [http://www.forbes.com/global/2001/0205/088\\_print.html](http://www.forbes.com/global/2001/0205/088_print.html)
4. Maybury, M., D'Amore, R., House, D. (2001) *Automated Discovery and Mapping of Expertise*. In Mark Ackerman, Andrew Cohen, Volmar Pipack, and Volker Wulf (eds.), *Beyond Knowledge Management: Sharing Expertise*, Cambridge: MIT Press
5. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. (1993) *Introduction to WordNet: An On-Line Lexical Database*, University of Princeton, U.S.A.
6. Sekine, S., Grishman, R. (1995) A Corpus-based Probabilistic Grammar with only Two Non-Terminals, In *Proceedings of the Fourth International Workshop on Parsing Technology*, pp.216-223
7. Tacit (1999) White paper of KnowledgeMail, <http://www.tacit.com/knowledge@mail>
8. Verschueren, J. (1980) *On Speech Act Verbs*, John Benjamins, Amsterdam
9. Vivacqua, A., Lieberman, H. (2000) Agents to Assist in Finding Help, In *Proceedings of the CHI2000*, pp.65-72

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**